UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/750,011	12/31/2003	Andrew S. Laucius	MS306413.1	9331

27195   7590   10/16/2008
AMIN, TUROCY & CALVIN, LLP
127 Public Square
57th Floor, Key Tower
CLEVELAND, OH 44114

EXAMINER
SCIACCA, SCOTT M

ART UNIT	PAPER NUMBER
2446	

NOTIFICATION DATE	DELIVERY MODE
10/16/2008	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docket1@thepatentattorneys.com
hholmes@thepatentattorneys.com
lpasterchek@thepatentattorneys.com

-- *The MAILING DATE of this communication appears on the cover sheet with the correspondence address* --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *06 August 2008*.

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-15 and 17-28* is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-15 and 17-28* is/are rejected.

7)☒ Claim(s) *23* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *31 December 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All   b)☐ Some * c)☐ None of:

       1.☐ Certified copies of the priority documents have been received.

       2.☐ Certified copies of the priority documents have been received in Application No. _____.

       3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

This office action is responsive to communications filed on August 6, 2008.

Claim 16 is cancelled. Claims 1, 19, 22, 24 and 26 are in independent form. Claims 1,

19, 22, 24 and 26 have been amended. Claims 1-15 and 17-28 are pending in the

application.

Continued Examination Under 37 CFR 1.114

1.   A request for continued examination under 37 CFR 1.114, including the fee set

forth in 37 CFR 1.17(e), was filed in this application after final rejection. Since this

application is eligible for continued examination under 37 CFR 1.114, and the fee set

forth in 37 CFR 1.17(e) has been timely paid, the finality of the previous Office action

has been withdrawn pursuant to 37 CFR 1.114. Applicant's submission filed on August

6, 2008 has been entered.

Claim Rejections - 35 USC § 101

2.   35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of
> matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the
> conditions and requirements of this title.

3.   Claims 1-15, 17, 18 and 24-28 are rejected under 35 U.S.C. 101 because the

claimed invention is directed to non-statutory subject matter.

Claim 1 is directed towards a "system that facilitates incremental web crawls."

The elements of the claimed system are "an indexer that places items with similar

properties into respective chunks" and "a chunk map." The indexer may be interpreted

as software routines since the specification does not explicitly state that the indexer

must be a physical machine. The chunk map is interpreted as a group of data. In order

for the claimed system to qualify as a machine under 35 U.S.C. 101, at least one of the

claimed elements must be a device or a physical part of a device. Software, per se and

data are not considered to be devices or parts of a device. Claims 2-15, 17 and 18 are

rejected accordingly because of their dependence from Claim 1.

Claim 24 is directed towards a "data packet." A packet by itself is not a process,

machine, manufacture, or a composition of matter because it is only considered to be a

collection of data. Claim 25 is rejected because of its dependence from Claim 24.

Claim 26 is directed towards a "system that facilitates increment web crawls."

The elements of the claimed system are "means for placing items with similar properties

into respective chunks" and "means for storing at least some of the properties", which

are considered to be the same as the "indexer" and "chunk map" from Claim 1,

respectively. The indexer may be interpreted as software routines since the

specification does not explicitly state that the indexer must be a physical machine. The

chunk map is interpreted as a group of data. In order for the claimed system to qualify

as a machine under 35 U.S.C. 101, at least one of the claimed elements must be a

device or a physical part of a device. Software, per se and data are not considered to

be devices or parts of a device. Claims 27 and 28 are rejected because of their

dependence from Claim 26.

Claim Rejections - 35 USC § 112

4.   The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly
> claiming the subject matter which the applicant regards as his invention.

5.   Claims 1-15, 17, 18, 22, 23 and 26-28 are rejected under 35 U.S.C. 112, second

paragraph, as being indefinite for failing to particularly point out and distinctly claim the

subject matter which applicant regards as the invention.

    Claim 1 recites the limitation "documents comprising a particular chunk" in lines

5-6. Claim 22 recites the limitation "documents comprising a particular chunk" in line 5.

Claim 26 recites the limitation "documents comprising a particular chunk" in lines 5-6.

There is insufficient antecedent basis for this limitation in the claims. Claims 2-15, 17

and 18 are rejected accordingly because of their dependence from Claim 1, Claim 23 is

rejected because of its dependence from Claim 22 and Claims 27 and 28 are rejected

because of their dependence from Claim 26.

Claim Rejections - 35 USC § 103

6.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set
> forth in section 102 of this title, if the differences between the subject matter sought to be patented and
> the prior art are such that the subject matter as a whole would have been obvious at the time the
> invention was made to a person having ordinary skill in the art to which said subject matter pertains.
> Patentability shall not be negatived by the manner in which the invention was made.

7.   Claims 1-15, 17, 18, 22, 23 and 26-28 are rejected under 35 U.S.C. 103(a) as

being unpatentable over Dean et al. (US 7,305,610) in view of Najork et al. (US

6,263,364).

Regarding Claim 1, Dean teaches a system that facilitates incremental web crawls comprising:

an indexer that places items with similar properties into respective chunks (*"The present invention provides innovative techniques for crawling of hyperlinked documents"* – See Col. 1, lines 41-42; *"The links to hyperlinked documents are grouped by host at a step 403"* – See Col. 7, lines 23-24; Links to hyperlinked documents having a common host are grouped together); and,

a chunk map that stores at least some of the properties associated with the respective chunk (*"In order to accomplish rate limiting of hosts, each host has an associated stall time, which is the earliest time at which another link from this host should be crawled or released to a crawler"* – See Col. 6, lines 47-50; The stall times are a property associated with a respective host), the stored properties are shared by all the items in the respective chunk (Since all the links (items) belonging to a particular host are grouped together and each host has an associated stall time (property), the stall time is shared by all links in the group), the chunk map employed to facilitate an incremental web re- crawl, wherein the properties of each chunk stored in the chunk map are utilized to determine a re-crawl of that chunk (*"At a step 405, a host to crawl next is selected according to a stall time of the host. The stall time can indicate the earliest time that the host should be crawled"* – See Col. 7, lines 25-27).

Dean does not explicitly teach that the properties are at least one of average time between change or average importance of documents comprising a particular chunk.

However, Najork discloses placing items with similar properties into respective chunks, where the properties are average time between change (*"the document is assigned to a priority level subqueue based on a predefined set of criteria 282 are satisfied, including but not limited to: ... the document's rate of change, based on (a) its modification date and time"* – See Col. 11, lines 48-55).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the system taught by Dean so that the stored properties include average rate of change of documents comprising a particular chunk. Motivation for doing so would be to have a mechanism for keeping the results of a crawl up to date, using a continuous crawl that is biased toward pages that are most likely to have been changed since the last time the crawler fetched them (See Najork, Col. 3, lines 51-55)

Regarding Claim 2, Dean teaches the items comprising information associated with a Uniform Resource Locator (*"a single link (e.g., uniform resource locator or URL)"* – See Col. 1, lines 46-47).

Regarding Claim 3, Dean teaches the items comprising at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file (*"content filters 205 can process the contents of the hyperlinked document according to the type of the file"* – See Col. 4, lines 34-35; *"For some file types (e.g., HTML pages, postscript files and PDF files), the canonical version of the file as it was extracted from the Web can be stored by store managers"* – See Col. 4, lines 38-40).

Regarding Claim 4, Dean teaches the items receives from a crawler, the crawler

responsible for a specific set of Uniform Resource Locators (*"crawlers 203 will*

periodically request new batches of links" – See Col. 6, lines 24-25; As mentioned

above, the links include URLs).

Regarding Claim 5, Dean teaches a master control process that can modify the

chunk map to facilitate load balancing amongst a plurality of crawlers (*"FIG. 7 shows a*

flow chart of a process of adjusting stall times" – See Col. 7, lines 34-35; *"Once the*

actual retrieval time is determined, the stall time for the selected host can be adjusted

according to the retrieval time at a step 503" – See Col. 7, lines 40-43; *"each computer*

system 1 can be executing one or more web crawler that traverses hyperlinked

documents and saves information regarding the traversed hyperlinked documents on

the computer system" – See Col. 3, lines 60-63).

Regarding Claim 6, Dean teaches a master control process that serves as an

interface between a crawler and a re-crawl controller (*"Crawlers 203 are responsible for*

retrieving hyperlinked documents from the servers" – See Col. 4, lines 20-21 *"A link*

(e.g., URL) server 201 determines which links should be crawled next" – See Col. 4,

lines 13-14; Fig. 4 shows an interface between the crawler and the re-crawl controller).

Regarding Claim 7, Dean teaches the master control process maintaining a

known chunks table that stores information for components of a system (*"link managers*

215 are responsible for keeping track of the status (the states described above) of each

link in the system" – See Col. 5, lines 34-36).

Regarding Claim 8, Dean teaches the master control process exposing an

interface for communication with a component of the system (Fig. 4 shows an interface

between the crawler and the re-crawl controller).

Regarding Claim 9, Dean teaches the interface returning a list of chunks the

component should have and where to get the chunks (*"Link server 201 maintains a pool*

of uncrawled links and groups the links by the host on which each link resides" – See

Col. 4, lines 14-16; *"When a crawler needs one or more links to crawl, the crawler*

requests one or more links from link server 201" – See Col. 4, lines 26-27).

Regarding Claim 10, Dean teaches the interface returning a list of the chunks

that should be actively served by the component (*"When a crawler needs one or more*

links to crawl, the crawler requests one or more links from link server 201" – See Col. 4,

lines 26-27).

Regarding Claim 11, Dean teaches the interface returning a range of chunk

identifiers to use in building a new chunk by the component (*"The host to crawl next is*

selected" – See Col. 2, lines 2-3).

Regarding Claim 12, Dean teaches the interface causing an old chunk to be

retired by the system (*"Once this host is found, a link is selected from the host, made*

ready to be passed to a crawler 203 and the link is removed from the hosts set of

uncrawled links" – See Col. 7, lines 7-10; Each chunk is a group of links sharing a

common host. Once all the links are crawled they are removed).

Regarding Claim 13, Dean teaches the master control process facilitating

movement of chunks from one component to another component (*"Link server 201*

maintains a pool of uncrawled links and groups the links by the host on which each link

resides" – See Col. 4, lines 14-16; *"When a crawler needs one or more links to crawl,*

the crawler requests one or more links from link server 201" – See Col. 4, lines 26-27).

Regarding Claim 14, Dean teaches movement of chunks being based, at least in

part, upon at least one of rebalancing index servers after one goes down, re-crawling

pages previously crawled, and, restoring a state of a crawler after it has crashed (*"Link*

server 201 maintains a pool of uncrawled links and groups the links by the host on

which each link resides" – See Col. 4, lines 14-16; *"When a crawler needs one or more*

links to crawl, the crawler requests one or more links from link server 201" – See Col. 4,

lines 26-27; *"each host has an associated stall time, which is the earliest time at which*

another link from this host should be crawled or released to a crawler" – See Col. 6,

lines 47-50; Chunks (groups of links sharing a common host) are assigned by a link

server to a crawler based on a stall time. The stall time determines how often links

should be re-crawled).

Regarding Claim 15, Dean teaches a re-crawl component that employs the

chunk map to determine which chunks, if any, to re-crawl at a particular time (*"each host*

has an associated stall time, which is the earliest time at which another link from this

host should be crawled or released to a crawler" – See Col. 6, lines 47-50; The stall

time indicates which chunk (group of links having a common host) should be re-crawled

at a particular time by specifying how long a crawler must wait before it may re-crawl the

links belonging to a particular host).

Regarding Claim 17, Dean teaches an index chunk that stores information

associated with an index of at least some of the items (*"there may be PageRank*

processes 219 that retrieve links from links files 217 and provides the links with a

priority or rank" – See Col. 5, lines 65-67).

Regarding Claim 18, Dean teaches a rank chunk that stores a static rank

associated with an index chunk (*"there may be PageRank processes 219 that retrieve*

links from links files 217 and provides the links with a priority or rank" – See Col. 5, lines 65-67).

Regarding Claim 22, Dean teaches a method of performing document re-crawl comprising:

accessing a chunk map containing properties associated with respective chunks of data as a result of one or more web crawls (*"The present invention provides innovative techniques for crawling of hyperlinked documents"* – See Col. 1, lines 41-42; *"The links to hyperlinked documents are grouped by host at a step 403"* – See Col. 7, lines 23-24; *"In order to accomplish rate limiting of hosts, each host has an associated stall time, which is the earliest time at which another link from this host should be crawled or released to a crawler"* – See Col. 6, lines 47-50), the stored properties are shared by all the items in the respective chunk (Since all the links (items) belonging to a particular host are grouped together and each host has an associated stall time (property), the stall time is shared by all links in the group); and,

periodically determining, based on the properties of each chunk in the chunk map, whether to re-crawl the chunk of data (*"At a step 405, a host to crawl next is selected according to a stall time of the host. The stall time can indicate the earliest time that the host should be crawled"* – See Col. 7, lines 25-27).

Dean does not explicitly teach that the properties are at least one of average time between change or average importance of documents comprising a particular chunk.

However, Najork discloses placing items with similar properties into respective chunks, where the properties are average time between change (*"the document is assigned to a priority level subqueue based on a predefined set of criteria 282 are satisfied, including but not limited to: ... the document's rate of change, based on (a) its modification date and time"* – See Col. 11, lines 48-55).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify the system taught by Dean so that the stored properties include average rate of change of documents comprising a particular chunk for the same reasons as those given with respect to Claim 1.

Regarding Claim 23, Najork teaches the period determination being based, at least in part, upon, at least one of average time between change and average importance of documents comprising a particular chunk (*"the document is assigned to a priority level subqueue based on a predefined set of criteria 282 are satisfied, including but not limited to: ... the document's rate of change, based on (a) its modification date and time"* – See Col. 11, lines 48-55).

Regarding Claim 26, Dean teaches a system that facilitates increment web crawls comprising:

means for placing items with similar properties into respective chunks (*"The present invention provides innovative techniques for crawling of hyperlinked documents"*

– See Col. 1, lines 41-42; *"Link server 201 maintains a pool of uncrawled links and groups the links by the host on which each link resides"* – See Col. 4, lines 14-16); and,

means for storing at least some of the properties associated with the respective chunk (*"In order to accomplish rate limiting of hosts, each host has an associated stall time, which is the earliest time at which another link from this host should be crawled or released to a crawler"* – See Col. 6, lines 47-50; *"By utilizing stall times, embodiments of the invention can ensure that hosts are not crawled too quickly. The stall times can be a predetermined amount of time, vary according to host and vary according to the actual response time of the host"* – See Col. 7, lines 31-34), wherein the stored properties are shared by all the items in the respective chunk (Since all the links (items) belonging to a particular host are grouped together and each host has an associated stall time (property), the stall time is shared by all links in the group), and

employing the stored properties of each chunk to facilitate an incremental web re-crawl (*"At a step 405, a host to crawl next is selected according to a stall time of the host. The stall time can indicate the earliest time that the host should be crawled"* – See Col. 7, lines 25-27).

Dean does not explicitly teach that the properties are at least one of average time between change or average importance of documents comprising a particular chunk.

However, Najork discloses placing items with similar properties into respective chunks, where the properties are average time between change (*"the document is assigned to a priority level subqueue based on a predefined set of criteria 282 are*

satisfied, including but not limited to: ... the document's rate of change, based on (a) its

modification date and time" – See Col. 11, lines 48-55).

It would have been obvious to one of ordinary skill in the art at the time the

invention was made to modify the system taught by Dean so that the stored properties

include average rate of change of documents comprising a particular chunk for the

same reasons as those given with respect to Claim 1.

Regarding Claim 27, Dean teaches the items comprising information associated

with a Uniform Resource Locator (*"a single link (e.g., uniform resource locator or URL)"*

– See Col. 1, lines 46-47).

Regarding Claim 28, Dean teaches the items comprising at least one of an HTML

file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file (*"content filters 205 can*

process the contents of the hyperlinked document according to the type of the file" –

See Col. 4, lines 34-35; *"For some file types (e.g., HTML pages, postscript files and*

PDF files), the canonical version of the file as it was extracted from the Web can be

stored by store managers" – See Col. 4, lines 38-40).

8.   Claims 19-21 are rejected under 35 U.S.C. 103(a) as being unpatentable over

Dean et al. (US 7,305,610) in view of Evans et al. (US 2004/0030683).

Regarding Claim 19, Dean teaches a method of performing document re-crawl comprising:

parsing a first chunk for uniform resource locators (*"The present invention provides innovative techniques for crawling of hyperlinked documents"* – See Col. 1, lines 41-42; *"At a step 401, links to hyperlinked documents are received. The links are to hyperlinked documents that are to be crawled. The links to hyperlinked documents are grouped by host at a step 403"* – See Col. 7, lines 21-24; *"a single link (e.g., uniform resource locator or URL)"* – See Col. 1, lines 46-47; *"A link (e.g., URL) server 201 determines which links should be crawled next. Link server 201 maintains a pool of uncrawled links and groups the links by the host on which each link resides"* – See Col. 4, lines 13-16; Links to hyperlinked documents having a common host are grouped together (in a chunk). A link server determines which links should be crawled next), wherein a chunk map that stores properties associated with the respective chunk stored in a chunk table is employed to determine the first chunk (*"In order to accomplish rate limiting of hosts, each host has an associated stall time, which is the earliest time at which another link from this host should be crawled or released to a crawler"* – See Col. 6, lines 47-50; The stall times are a property associated with a respective host. The stall time is used to determine which host, and thus, which group of links (chunk) should be crawled next), wherein the stored properties are shared by all the items in the respective chunk (Since all the links (items) belonging to a particular host are grouped together and each host has an associated stall time (property), the stall time is shared by all links in the group), and

re-crawling the uniform resource locators (*"Once the host to be crawled next is selected, a hyperlinked document from the selected host is crawled at a step 407"* – See Col. 7, lines 27-29).

Dean does not explicitly teach forming a second chunk separate from the first chunk, based at least in part, upon the re-crawled uniform resource locators.

However, Dean does mention that web pages may include URLs which point to another web page on another host (*"In a wide area network such as the Internet, some of the computer systems are servers (or hosts)"* – See Col. 3, lines 39-40; *"The web pages typically include links in the form of uniform resource locators (URLs) that are a link to another web page, whether it is on the same server or a different one"* – See Col. 3, lines 44-47). Thus, in the process of crawling a web page of a first host, a crawler may find a link to a web page hosted by a second host.

Evans teaches a crawler that, upon encountering a new web site, will perform an exhaustive search of the site (*"the first time a web site (or any location of content, such as a file directory) is encountered, an exhaustive search is conducted"* – See [0023]).

Combining these features would yield a crawler that, upon encountering a link to a web page on a second host, would also perform a crawl of the web page on the newly discovered second host. This would also result in the formation of a second chunk, since the links encountered on the new host would be grouped together in a new chunk.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to modify Dean's method of performing web crawls to include upon encountering a link to a web page on a second host, performing a crawl of the web

page on the newly discovered second host. According to Evans, new content is

continuously appearing on the web (*"The volume and variety of informational content*

available on the web is likely to continue to increase at a rather substantial pace" – See

[0002]). It is the job of a crawler to index the content of the web so that search engines

may inspect the index and return search results to a user based on a search query (See

[0004]). Thus, performing a crawl of a web page on a newly discovered host (and

subsequently indexing the contents of the web page) would ensure that content which is

new to the web will be available to search engines.

Regarding Claim 20, Dean teaches moving the first chunk (*"When a crawler*

needs one or more links to crawl, the crawler requests one or more links from link

server 201" – See Col. 4, lines 26-27).

Regarding Claim 21, Dean teaches one or more computer readable media

having stored thereon computer executable instructions for carrying out the method of

claim 19 (*"FIG. 1 illustrates an example of a computer system that can be used to*

execute the software of an embodiment of the invention" – See Col. 2, lines 65-67;

"Although CD-ROM 15 is shown as an exemplary computer readable storage medium,

other computer readable storage media including floppy disk, tape, flash memory,

system memory, and hard drive can be utilized. Additionally, a data signal embodied in

a carrier wave (e.g., in a network including the Internet) can be the computer readable

storage medium" – See Col. 3, lines 8-14).

9.   Claims 24 and 25 are rejected under 35 U.S.C. 103(a) as being unpatentable

over Dean et al. (US 7,305,610) in view of Dingsor et al. (US 7,058,727).

Regarding Claim 24, Dean teaches a chunk header that includes metadata, the

metadata shared by all the items in the chunk (*"In order to accomplish rate limiting of*

hosts, each host has an associated stall time, which is the earliest time at which another

link from this host should be crawled or released to a crawler" – See Col. 6, lines 47-50;

Since all the links (items) belonging to a particular host are grouped together and each

host has an associated stall time (property), the stall time is shared by all links in the

group); and

document files that include content found on the Internet (*"content filters 205 can*

process the contents of the hyperlinked document according to the type of the file" –

See Col. 4, lines 34-35; *"For some file types (e.g., HTML pages, postscript files and*

PDF files), the canonical version of the file as it was extracted from the Web can be

stored by store managers" – See Col. 4, lines 38-40),

wherein the average of the at least one of the properties of all the document files

determines if the document should be re-crawled (*"At a step 405, a host to crawl next is*

selected according to a stall time of the host. The stall time can indicate the earliest time

that the host should be crawled" – See Col. 7, lines 25-27).

Although Dean mentions crawlers using the metadata and receiving document

files from the Internet (*"a data signal embodied in a carrier wave (e.g., in a network*

including the Internet) can be the computer readable storage medium" – See Col. 3,

lines 11-14), Dean does not explicitly teach assembling the above mentioned data into a

data packet.

Dingsor teaches an IP packet (See Fig. 5) having metadata (*"The first six words*

in the sequence are the IP header 544" – See Col. 6, lines 29-30) and data (*"the*

remaining words are in IP data area 546" – See Col. 6, lines 30-31). Dingsor mentions

that the Internet makes uses the TCP/IP protocol (*"the term 'Internet' refers to the*

collection of networks and gateways that use the TCP/IP suite of protocols" – See Col.

1, lines 24-26). Dingsor also teaches the packet having an offset section that provides

information associated with the data (*"Other fields in the IP header, like total length and*

fragment offset, are used to breakup network datagrams into packets at the source

computer and reassemble them at the destination computer" – See Col. 6, lines 44-47).

In order for two computers to communicate it is necessary to place the

information in a data packet, such as the IP packet taught by Dingsor. In order for

Dean's web crawler to function it is necessary in some steps to transmit data from one

computer to another (*"Links to hyperlinked documents to be crawled are stored and*

when it is determined that more links are desired, requests are sent to multiple link

managers for more links. Additional links are received from the link managers" – See

Col. 2, lines 12-15; *"FIG. 5 shows an example of a block diagram of a single link server*

receiving links from multiple link managers and storing the links in buckets grouped by

host" – See Col. 2, lines 36-38; *"FIG. 6 shows a flow chart of a process of crawling*

hyperlinked documents that includes selecting the host to crawl next according to a stall

time of the host" – See Col. 2, lines 39-41). It would have been obvious to one of ordinary skill in the art at the time the invention was made to assemble the metadata and document files taught by Dean along with the offset information taught by Dingsor into a data packet. Motivation for doing so would be to allow the information to be transmitted from one computer to another.

Regarding Claim 25, Dean teaches at least one of the document files comprising at least one of an HTML file, a PDF file, a PS file, a PPT file, an XLS file and a DOC file (*"content filters 205 can process the contents of the hyperlinked document according to the type of the file"* – See Col. 4, lines 34-35; *"For some file types (e.g., HTML pages, postscript files and PDF files), the canonical version of the file as it was extracted from the Web can be stored by store managers"* – See Col. 4, lines 38-40).

Response to Arguments

10.   Applicant's arguments with respect to Claims 1, 19, 22, 24 and 26 have been considered but are moot in view of the new grounds of rejection.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Scott M. Sciacca whose telephone number is (571) 270-1919. The examiner can normally be reached on Monday thru Friday, 7:30 A.M. - 5:00 P.M. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Jeff Pwu can be reached on (571) 272-6798. The fax phone number for the

organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the

Patent Application Information Retrieval (PAIR) system. Status information for

published applications may be obtained from either Private PAIR or Public PAIR.

Status information for unpublished applications is available through Private PAIR only.

For more information about the PAIR system, see http://pair-direct.uspto.gov. Should

you have questions on access to the Private PAIR system, contact the Electronic

Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a

USPTO Customer Service Representative or access to the automated information

system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Scott M. Sciacca/
Examiner, Art Unit 2446

/Jeffrey Pwu/
Supervisory Patent Examiner, Art Unit 2446